

DEQ's review of public comments submitted by Dominion, Virginia Manufactures Association (VMA), Mead Westvaco, and the Virginia Association of Municipal Wastewater Agencies (VAMWA)

The following questions were derived solely from the public comments submitted by the mentioned parties for the purpose of AAC and DEQ review.

- 1) Is it likely that a single metric in the index could incorrectly skew a site's characterization? (For example, the HBI could receive a 'good' score due to the presence and dominance of acidophilic stoneflies in a stream impaired by acid mine drainage, therefore, significantly influencing the final SCI score to the point of assessing the stream as "non-impaired")**

In general, each metric is relatively robust, such that the presence of "outlier" organisms or similar problems will not skew the metric to the point of leading to an incorrect conclusion. Also, when all of the bioassessment metrics are taken together into one bioassessment score, there is an additive effect of their robustness such that it becomes even more difficult to derive a misleading conclusion.

- 2) Since the SCI was predicated on detrital based communities, does the SCI have the ability to differentiate between acceptable and unacceptable benthic communities in seston rich tailwaters (waters below impoundments)?**

The biota in tailwaters certainly will be different than from a stream not affected by water release and the SCI likely would reflect that. How tailwaters are dealt is currently being investigated by DEQ staff.

- 3) Due to the complexity of aquatic systems, will it be possible for the SCI alone to be a reliable and accurate assessment tool? Will it be necessary to include other biological attributes such as using fish and habitat metrics to take in account background variability?**

The use of macroinvertebrates alone has become widely accepted as one way to assess stream health. Use of other attributes, such as fish and habitat, could also be used in conjunction with the SCI; however, the SCI has the ability to stand alone.

- 4) In the development of the SCI, Tetra Tech identified variation in the database that was 10% of the mean. Is this sufficient variation such that a site could have a 'failing' score when in fact it only represented confounding background information?**

The recommendation of Tetra Tech is that the 10th percentile of SCI for reference sites be used as the criteria. The value is not 10% of the mean but the value in the reference data that is smaller than 90% of the reference values. This implies that 10% of samples that are actual reference sites will be classified as non-reference. It is common in a decision process that errors will be made and the 10% essentially sets the error rate for the

reference set. The misclassified sites may be sites that represent confounding background information. Metric-based bioassessment tools typically take into account the large degree of variability that can occur and set the lower threshold of acceptability based on that variation.

5) Many habitat related metrics (e.g., epifaunal substrate score) can experience a great deal of site-to-site variability that is not necessarily related to anthropogenic stress. Therefore, is it likely that many sites will not score well on the SCI simply because the natural habitat is not as favorable as at reference sites?

This is part of the natural variability in the metric scores, dealt with in the question above.

6) Does the SCI have the ability to address ‘cause and effect’ in an evaluation of site differences (reference vs. test sites)? Is it necessary to establish a cause and effect relationship in order for the SCI to be an effective assessment tool?

The SCI does not necessarily address cause and effect, though it may be able to do so under certain conditions. It simply notes that a stream reach is or is not impaired relative to the reference condition.

7) Is the SCI appropriate for streams greater than 4th order?

An SCI approach is appropriate for streams >4th order. However, the VSCI developed by Tetra Tech for Virginia would have to be tested before being considered appropriate since the data upon which it was based did not include >4th order streams

8) The setting of the reference condition at the 10th percentile of reference sites assumes that 10% of the reference sites were impaired by unknown causes which demonstrate that even non-stressed sites may experience low SCI scores. Should the threshold of impairment definition be evaluated using both the reference site distribution and the stressed site distribution in order to allow a calculation of a threshold as a compromise between the probability of Type I and Type II errors?

The model that is used implies a Type I error rate of 10%. If a new site is randomly selected from the reference population, it is expected that only 10% will fall below the threshold. No evaluation of Type II errors is given although it would be valuable to test the procedure for biological evaluation.

9) Is the use of a single reference condition for all streams in five ecoregions and multiple stream orders appropriate, especially considering the lack of information for the Central Appalachians and an insufficient number of reference sites for the Piedmont?

The conclusion of the AAC was that the VSCI was an appropriate assessment tool for the Blue Ridge Mountains, Northern Piedmont and the Central Appalachian Ridges and Valleys ecoregions. More data needs to be collected in the Central Appalachians and

Piedmont ecoregions in order to determine if separate reference conditions are needed. DEQ biologists are currently collecting more data in these ecoregions.

10) Could setting the standard (best value) of the 95th percentile for use in scoring each metric artificially lower a site's score?

No.

11) The results of the MDS analysis were based on visual examination of scatter plots. Does the fact that no statistical tests were performed for the assessment and groupings undermine the scientific basis of the SCI? Should Permutation or Monte Carlo testing for confounding factors be used to test for significant differences?

The point of the graphical displays is to suggest separation or to indicate where differences may be. It is certainly possible to test for differences in ecoregion or other factors using the NMDS scores. It is also possible to carry out the tests on the raw data rather than the scores. NMDS is intended to be descriptive in the sense of summarizing the distances between the samples. As a separation method it may be inefficient unless the distances are primarily associated with separation. DEQ is currently evaluating statistical tests on the data as a part of the SCI recalibration/validation process.

12) Does family level macroinvertebrate taxonomy provide sufficient resolution for accurate functional feeding group classification?

Various studies have shown that family taxonomy is sufficient for this purpose, although others have argued that genus level taxonomy is required. However, Tetra Tech found that % scrapers at the family level had discriminatory power between reference and stressed sites.

13) Are % scrapers and % Chironomidae appropriate metrics for distinguishing between impacted and non-impacted sites?

They are two metrics that are widely regarded as being able to discriminate between impaired and non-impaired sites. Tetra Tech found that they had discriminatory power in this geographic region and a VCU study did the same.

14) Is it possible that the high degree of similarity observed between the two sampling seasons may be because the time frames assigned to the seasons do not represent two seasons but rather three or possibly four seasons with the potential of a fair amount of faunal overlap? To limit variability, is a single well defined index period needed?

The narrower the sampling period window the better in terms of reducing variability; however, the seasonal cycles of most organisms are such that they cover long time periods, thus negating the need for much concern for this type of error. In general,

benthic invertebrates have their greatest differences in terms of community composition between “summer” and “winter” communities.

15) Was the HBI for the SCI based on macroinvertebrate tolerance data from Wisconsin? If this is the case, would the HBI metric introduce flaws into the index?

No. Macroinvertebrate tolerance data for Virginia were either supplied by DEQ biologists or were determined by consulting Barbour et al. 1999, Merrit and Cummings 1996, or best professional judgment.

16) Should there be more effort for understanding how the SCI performs with varying degrees of impairment (along a gradient)?

This would be nice to know, but is not mandatory for the development and use of the index.

17) Since different methods of data gathering (sampling) were used in different VADEQ Regions, was it possible for Tetra Tech investigators to provide an accurate analysis of ecoregional and VADEQ Region differences and was it possible that it significantly affected their ability to estimate index variability due to measurement error?

The extent of the differences in sampling were considered minimal to the point where variation should not have affected the index development according to the Academic Advisory Committee.

18) Would having the DEQ attempt to determine if different stressors have different effects on the individual biological metrics and the SCI itself improve the accuracy of the SCI?

The SCI is a tool to assess the cumulative effects of all stressors. It can assess a single stressor if it is known that only one stressor is the problem, but it can not tease out difference among stressors. Since the index is focused on cumulative effects, there is no reason to conduct tests on individual stressors unless DEQ requires an assessment tool for specific stressors.

19) Not including habitat metrics, should only non-biological factors be used in determining reference sites in order to avoid circular logic in the SCI development process?

Only non-biological data were used determine reference sites.

20) Psuedoreplication could have the effect of underestimating the variance of sample data, which could in turn cause an overestimation of the apparent differences between sample data subsets. Could the psuedoreplication issue

therefore cause an overestimation of the ability of the SCI to distinguish between impaired and reference sites?

Yes, and thus pseudoreplication should be kept to a minimum or, better, yet, eliminated. Tetra Tech understood this problem and dealt with it as best that they could. DEQ staff will address the pseudoreplication issue before the SCI is implemented.

21) Should stream site assessment using the SCI be done using multiple samples do to within site variability?

The sampling method is one that essentially takes multiple samples but then pools the samples. It is not known if this has been tested elsewhere (i.e., pooled vs. multiple samples), but the pooled sampling method is standard protocol.

22) In its current form, can the SCI make accurate stream quality assessments or are there still significant issues with data quality and/or quantity?

There are issues that need to be addressed, but overall the SCI can work appropriately. Data quality and quantity issues are currently being addressed by DEQ staff.